Results of SemTab 2023

Oktie Hassanzadeh[1,*], Nora Abdelmageed[2], Vasilis Efthymiou[3], Jiaoyan Chen[4], Vincenzo Cutrona[5], Madelon Hulsebos[6], Ernesto Jiménez-Ruiz[7,8], Aamod Khatiwada[9], Keti Korini[10], Benno Kruit[11], Juan Sequeda[12] and Kavitha Srinivas[1]

[1] *IBM Research, USA*
[2] *Friedrich Schiller University Jena, Germany*
[3] *FORTH-ICS, Greece*
[4] *University of Manchester, UK*
[5] *SUPSI, Switzerland*
[6] *University of Amsterdam, The Netherlands*
[7] *City, University of London, UK*
[8] *SIRIUS, University of Oslo, Norway*
[9] *Northeastern University, USA*
[10] *University of Mannheim, Germany*
[11] *Vrije Universiteit Amsterdam, The Netherlands*
[12] *data.world, USA*

Abstract
SemTab 2023 was the fifth edition of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, collocated with the 22nd International Semantic Web Conference (ISWC) and the 18th Ontology Matching (OM) Workshop. SemTab provides a framework to conduct a systematic evaluation of state-of-the-art semantic table interpretation systems. In this paper, we give an overview of the 2023 edition of the challenge and summarize the results.

Keywords
Tabular data, Knowledge Graphs, Matching, SemTab Challenge, Semantic Table Interpretation

1. Motivation

Tabular data is prevalent on the Web and in enterprise data lakes, data catalogs, and other data repositories, and is often the primary data format used in data science and data analytics solutions. In practice, there is often a wide gap between the producers and consumers of tabular data. Data producers have the primary role of storage, maintenance, and availability of the raw data and

SemTab'23: Semantic Web Challenge on Tabular Data to Knowledge Graph Matching 2023, co-located with the 22nd International Semantic Web Conference (ISWC), November 6-10, 2023, Athens, Greece
*Corresponding author.
hassanzadeh@us.ibm.com (O. Hassanzadeh); nora.abdelmageed@uni-jena.de (N. Abdelmageed); vefthym@ics.forth.gr (V. Efthymiou); jiaoyan.chen@cs.ox.ac.uk (J. Chen); vincenzo.cutrona@supsi.ch (V. Cutrona); m.hulsebos@uva.nl (M. Hulsebos); ernesto.jimenez-ruiz@city.ac.uk (E. Jiménez-Ruiz); khatiwada.a@northeastern.edu (A. Khatiwada); kkorini@uni-mannheim.de (K. Korini); b.b.kruit@vu.nl (B. Kruit); juan@data.world (J. Sequeda); kavitha.srinivas@ibm.com (K. Srinivas)

often share the data without much metadata or with metadata in non-standard or textual form. On the other hand, data consumers need to identify the data they require, select relevant portions of the data, and refine and integrate the raw data to make the data usable in their application. This process of making the data consumable is often not feasible without the aid of automated solutions. A key enabler of automated solutions is the annotation of data elements with entities, classes, and relations in a knowledge graph (KG). Such annotations enable knowledge-based data discovery [1, 2, 3, 4], organization [5], integration [6, 7], and augmentation [8]. Automating the process of matching tabular data with KGs, also referred to as Semantic Table Interpretation (STI), has been the topic of extensive research in the literature [9, 10, 11, 12, 13, 14, 15, 16, 17, 18].

The Semantic Web Challenge on Tabular Data to Knowledge Graph Matching (SemTab) started in 2019 with the goal of providing an avenue for benchmarking and evaluation of various STI solutions. Over the years, the SemTab participants have proposed a range of solutions incorporating a variety of approaches to automated matching, with their key strengths and weaknesses analyzed using different datasets and rounds of each of the SemTab editions. In this paper, we provide a high-level summary of the 2023 edition of the SemTab challenge, along with the results.

2. The Challenge

The SemTab 2023 challenge included two tracks: the **Accuracy Track**, which is the standard track for evaluation of the accuracy of the solutions; and the **Datasets Track**, which focuses on new datasets and applications. The datasets track was also open to the submission of revisions of the existing datasets. As in 2022, SemTab 2023 also featured an *Artifacts Availability Badge*.

2.1. Accuracy Track

The Accuracy Track included 2 rounds, running from April 14 to June 22, 2023. The different rounds of SemTab 2023 have been organised to evaluate participating systems on various datasets, tasks, and target KGs, with variable difficulty. As with last year and unlike the initial editions of the challenge, where the rounds were run with the support of AIcrowd, we asked the participants to submit their solutions using a submission form, and the outcome was evaluated at the end of each round.

2.1.1. Datasets

The different datasets used to run SemTab 2023 rounds are reported in Table 1, with some statistics available in Tables 2 and 3. As with last year and unlike the initial editions where the ground truth was hidden from the participants, we provided partial ground truth data to the participants during the challenge in the form of a training and/or validation set. The teams could use these ground-truth labels to evaluate their methods locally. All the datasets are available in Zenodo. We used four groups of datasets across the two rounds:

- **WikidataTables**: datasets with tables generated using an improved version of our data generator that creates realistic-looking tables using SPARQL queries [19]. The target KG

for this dataset is Wikidata, and as with previous years the tasks are Cell Entity Annotation (CEA), Column Type Annotation (CTA), and Column Property Annotation (CPA). As reported in Table 2, the test consists of 9,917 tables, with an average of 2.5 columns and 4.7 rows. The dataset was generated using a configuration that resulted in a large number of very small tables, with a high level of ambiguity for entity columns. This was done by filtering for labels that can refer to more than one entity in Wikidata.
Link: https://doi.org/10.5281/zenodo.8393535

- **tFood [20]**: a dataset for tabular data to knowledge graph matching. It is derived for the Food domain and has two types of tables: 1) “horizontal" relational tables where each table represents a collection of entities, and 2) “entity" tables, each representing a single entity. We provided ground truth mappings to Wikidata for CEA, CTA, and CPA tasks in addition to a new Topic Detection (TD) task that aims at annotating an entire table to instances/entities or types/classes. The test set contains 3,945 horizontal and 7,643 entity tables. The horizontal tables have on average 5.5 columns and 19.6 rows, while the entity tables have on average 3.9 rows.
  Link: https://doi.org/10.5281/zenodo.7828163
- **SOTAB [21]**: a benchmark dataset created using tables from the WDC Schema.org Table Corpus for the CTA and CPA tasks used in both challenge rounds of SemTab. The datasets for SemTab were created by downsampling the original benchmark, with the intention of having easier CTA and CPA tasks to solve in the first round and harder tasks in the second round of the challenge. Therefore, the datasets of the first round have a smaller vocabulary corresponding to more general labels, while the datasets of the second round contain larger vocabularies with more specific labels. The column types and the column relationships are annotated using the Schema.org and DBpedia vocabularies and all the ground truths of all test sets are manually verified.
  Link: https://doi.org/10.5281/zenodo.8422037
- **CQA [22]**: a dataset based on Wikary [23], consisting of Wikipedia tables with ground truth annotation of both main properties and qualifiers from Wikidata for n-ary relations that are expressed by three table columns. Tables that had an overlap with multiple statements from Wikidata were selected, after which matching rows were removed from the tables. Participants were presented with example tables from the Simple English Wikipedia edition, and evaluated on 844 tables from the standard English edition. An example is shown in Table 4. The occurrence of qualifiers is highly skewed, and thus presents a challenge with regard to class imbalance.
  Link: https://doi.org/10.5281/zenodo.8398347

2.1.2. Evaluation measures

As per the previous editions, systems have been evaluated on a single annotation for each provided target for all the tasks. This means that in CEA, target cells are to be annotated with a single entity from the target KG. In CPA the target column pairs are to be annotated with a single property. In CTA, target columns are to be annotated with a single type from the target KG, which should be

Table 1
Datasets used across SemTab 2023 rounds.

	Rounds		Tasks					Target KGs		
	R1	R2	CTA	CPA	CEA	TD	CQA	DBpedia	Wikidata	Schema.org
WikidataTables	✓		✓	✓	✓				✓	
tFood	✓		✓	✓	✓	✓			✓	
SOTAB	✓	✓	✓	✓				✓		✓
CQA		✓					✓		✓	

Table 2
Statistics of the WikidataTables, tFood, and CQA datasets in each SemTab 2023 round.

Validation	WikidataTables	tFood (horizontal)	tFood (entity)	CQA
Tables #	500	438	849	843
Avg. # Cols	2.46	6.24	2.00	5.72
Avg. # Rows	5.95	30.04	4.05	44.47
Test	**WikidataTables**	**tFood (horizontal)**	**tFood (entity)**	**CQA**
Tables #	9,917	3,945	7,643	844
Avg. # Cols	2.51	5.51	2.00	5.91
Avg. # Rows	4.66	19.63	3.86	44.18

Table 3
Statistics of the SOTAB datasets in each SemTab 2023 round.

	SOTAB-CTA		SOTAB-CPA	
	Round 1	Round 2	Round 1	Round 2
Tables #	26,271	44,407	17,191	28,222
Avg. # Cols	7.48	8.16	8.60	9.48
Avg. # Rows	178.07	177.56	196.57	224.67

Table 4
Example table for CQA task, about Oscar nominations. The main property "nominated for" (P1411) holds between columns 2 and 0 (the subject and object column), which is expanded upon with the qualifier "for work" (P1686) in column 3 (the qualifier column). This property-qualifier pair describes the n-ary relation that holds between these three columns.

0	1	2	3	4
Academy Award for Best Actor	2000(73rd)	Russell Crowe	Gladiator	Maximus Decimus Meridius
Academy Award for Best Actor	2000(73rd)	Javier Bardem	Before Night Falls	Reinaldo Arenas
Academy Award for Best Actor	2000(73rd)	Ed Harris	Pollock	Jackson Pollock

as fine-grained as possible, i.e., the most specific type or the lowest applicable type in the type hierarchy. Similarly, TD and CQA tasks required a single annotation in the output.

The evaluation measures for CEA, CPA and CTA are the standard Precision, Recall and F1-score, as defined in Equation 1:

$$P = \frac{|\text{Correct Annotations}|}{|\text{System Annotations}|},\ R = \frac{|\text{Correct Annotations}|}{|\text{Target Annotations}|},\ F1 = \frac{2 \times P \times R}{P + R} \tag{1}$$

where target annotations are the target cells for CEA, the target columns for CTA, and the target column pairs for CPA. We consider an annotation as *correct* if it is in the ground truth set. A target cell may have multiple annotations in the ground truth, because of redirect and same-as links in KGs.

As in the past editions, given the fine-grained type hierarchy in Wikidata, we used a modified notion of Precision and Recall in the CTA evaluation [24]. We adapt the numerators to consider partially correct annotations, *i.e.*, annotations that are either ancestors or descendants of the ground truth (GT) classes. The correctness score $cscore$ of a CTA annotation α takes into account the distance between the annotation and the GT classes in the type hierarchy, and it is defined as:

$$\text{cscore}(\alpha) = \begin{cases} 0.8^{d(\alpha)}, & \text{if } \alpha \text{ is in GT, or an ancestor of the GT, with } d(\alpha) \leq 5 \\ 0.7^{d(\alpha)}, & \text{if } \alpha \text{ is a descendant of the GT, with } d(\alpha) \leq 3 \\ 0, & \text{otherwise;} \end{cases} \quad (2)$$

where $d(\alpha)$ is the shortest distance to one of the ground truth classes. As for the CEA, CTA ground truth columns may have multiple classes. For example, $d(\alpha) = 0$ if α is a class in the GT (cscore$(\alpha) = 1$), and $d(\alpha) = 2$ if α is a grandchild of a class in the GT (cscore$(\alpha) = 0.49$). We do not consider types in the higher level(s) of the KG type hierarchy, *e.g.*, `Q35120 [entity]` in Wikidata. Given the correctness score $cscore$, the approximated Precision (AP), Recall (AR), and F1-score (AF1) for the CTA evaluation are calculated as follows:

$$AP = \frac{\sum cscore(\alpha)}{|\text{System Annotations}|}, \ AR = \frac{\sum cscore(\alpha)}{|\text{Target Annotations}|}, \ AF1 = \frac{2 \times AP \times AR}{AP + AR} \quad (3)$$

Finaly, CQA results are simply evaluated based on their accuracy. That is, the score for CQA is the number of correct property-qualifier pairs in the output divided by the total number of pairs in the ground truth. For convenience of display, we show the CQA accuracy scores in the Precision column in Table 7.

2.2. Datasets Track

The data that table-to-Knowledge-Graph matching systems are trained and evaluated on is critical for their accuracy and relevance. We invited dataset submissions that provide challenging and accessible new datasets to advance the state-of-the-art of table-to-KG matching systems. We encouraged datasets that provide tables along with their ground truth annotations for at least one of CEA, CTA and CPA tasks. The datasets could be general or specific to a certain domain.

Submissions were evaluated according to provide the following:

- Description of the data collection, curation, and annotation processes.
- Availability of documentation with insights in the dataset content.
- Publicly accessible link to the dataset (e.g. Zenodo) and its DOI.
- Explanation of maintenance and long-term availability.
- Clear description of the envisioned use-cases.
- Application in which the dataset is used to solve an exemplar task.

Table 5
Participation in the SemTab 2023 challenge.

	Wikidata Tables Round 1	tFood horizontal Round 1	tFood entity Round 1	SOTAB Schema.org Round 1	SOTAB Schema.org Round 2	SOTAB DBpedia Round 2	CQA Round 2
CEA	5	1	1	-	-	-	-
CTA	5	2	-	1	6	6	-
CPA	5	2	-	1	5	6	-
TD	-	2	2	-	-	-	-
CQA	-	-	-	-	-	-	3

Table 6
Round 1 Results

			TSOTSA		Kepler-aSI		TorchicTab		MUT2KG		Semtex	
Benchmark	**Target**	**Task**	**F1**	**Pr**	**F1**	**Pr**	**F1**	**Pr**	**F1**	**Pr**	**F1**	**Pr**
tFood horizontal	**Wikidata**	**TD**	0.013	0.013	**0.513**	0.513	-	-	-	-	-	-
		CEA	0.045	0.041	-	-	-	-	-	-	-	-
		CTA	0.031	0.031	**0.105**	0.105	-	-	-	-	-	-
		CPA	**0.285**	0.285	0.000	0.000	-	-	-	-	-	-
tFood entity	**Wikidata**	**TD**	**0.156**	1.000	0.000	0.000	-	-	-	-	-	-
		CEA	**0.237**	0.347	-	-	-	-	-	-	-	-
Wikidata Tables	**Wikidata**	**CEA**	0.627	0.627	0.006	0.959	0.830	0.839	0.408	0.587	**0.885**	0.904
		CTA	0.738	0.738	0.739	0.739	0.817	0.749	0.459	0.655	**0.934**	0.934
		CPA	0.102	0.102	0.777	0.777	0.934	0.934	0.226	0.948	**0.964**	0.968
SOTAB	**Schema.org**	**CTA**	-	-	**0.364**	0.343	-	-	-	-	-	-
		CPA	-	-	**0.235**	0.230	-	-	-	-	-	-

3. Results

Table 5 shows the participation for each dataset and round. Overall, we had 8 participants submitting to at least one round. We concluded Round 2 with 6 core participants, which submitted a system paper to the challenge: *Kepler-aSI* [25], *MUT2KG* [26], *DREIFLUSS* [27], *TSOTSA* [28], *SemTex* [29], and *TorchicTab* [30]. Most participants targeted SOTAB datasets (6 participants) and Wikidata Tables (5 participants), followed by the new CQA task (3 participants) and the tFood dataset (2 participants).

3.1. Accuracy Track Results

Tables 6 and 7 show the average F1-score and Precision/Accuracy achieved by the participating systems in Rounds 1 and 2, respectively. Unlike previous years that we had the top-performing systems performing well on all tasks and datasets, this year there is no single system that performs best across most or all the tasks. This is likely due to the newly introduced datasets and tasks that require addressing new and different kinds of challenges. In some ways this is similar to the Ontology Alignment Evaluation Initiative (OAEI) campaigns where most top-performing systems target only one or a small number of tracks.

Wikidata Tables. The top-performing system over Wikidata Tables on all three CEA, CTA, and

Table 7
Round 2 Results

			TSOTSA		Anu		Kepler-aSI		TorchicTab		MUT2KG		DREIFLUSS	
Benchmark	**Target**	**Task**	**F1**	**Pr**	**F1**	**Pr**	**F1**	**Pr**	**F1**	**Pr**	**F1**	**Pr**	**F1**	**Pr**
		CTA	0.3705	0.5601	0.5872	0.7048	0.3372	0.3601	**0.8966**	**0.8980**	0.3201	0.7922	0.3804	0.5767
SOTAB	**Schema.org**	**CPA**	0.2355	0.4339	0.6227	0.7915	-	-	**0.8711**	**0.8800**	0.7935	0.8483	0.1739	0.3196
		CTA	0.3906	0.6150	0.5330	0.6852	0.4611	0.4983	**0.8972**	**0.9093**	0.3333	0.8196	0.4097	0.6114
SOTAB	**DBpedia**	**CPA**	0.3186	0.4697	0.7212	0.8515	0.1316	0.132	**0.9049**	**0.9068**	0.823	0.8553	0.2069	0.3967
CQA	**Wikidata**	**CQA**	-	-	-	0.078	-	-	-	0.822	-	**0.872**	-	-

CPA tasks was the *SemTex* system, with F1 score of 0.885 on CEA, 0.934 on CTA, and 0.964 on CPA. In comparison, last year's top-performing system, *DAGOBAH* [31] achieved F1 scores of over 0.95 on all tasks. It would be interesting to evaluate *DAGOBAH* on this dataset and analyze whether the dataset this year is more difficult than previous years. Still, we expected this dataset to be the easiest among the datasets this year, which was confirmed by the results.

SOTAB. In Round 1, only *Kepler-aSI* system submitted results for SOTAB, with the highest F1 score of 0.36 for CTA and 0.24 for CPA, much lower than the system's performance on Wikidata Tables dataset, showing that SOTAB is an inherently different and more challenging dataset. In Round 2, five more systems participated on this dataset, with the *TorchicTab* system outperforming other systems, in some cases with a very high margin, with F1 scores of up to 0.9. This could be attributed to the use of pre-trained language models [32] that have shown very promising performance especially when dealing with Web data. It would be interesting to evaluate the performance of such solutions over domain-specific datasets with contents not derived from Web data.

tFood. TSOTSA and *Kepler-aSI* provided the highest numbers in Round 1, although the dataset proved to be very challenging for both systems. *TorchicTab* did not participate in Round 1 with this dataset, but the authors report good performance over the training set [30].

CQA. The best performance over CQA was achieved by the *MUT2KG* system, with an accuracy of 0.872, followed by *TorchicTab* with an accuracy of 0.822. These results prove the effectiveness of MUT2KG's neuro-symbolic approach to annotation [26].

3.2. Datasets Track Results

This year we received two submissions, out of which one was accepted for presentation at SemTab 2023: *TSOTSATable Dataset* [33]. The TSOTSATable Dataset presents an additional contribution to the food composition tables (FCTables) [34], introduced in the datasets track of SemTab 2022, containing tables that describe foods and their composition. The authors demonstrate their effort by annotating tabular data benchmarks for Food composition with Wikidata, FoodOn, and Open Research Knowledge Graph (ORKG) vocabularies. Interestingly, during the annotation, the authors found many tables to be irrelevant to the food domain. This led to the addition of a new annotation task, that they called Irrelevant Table Detection (ITD). The goal of this task is to detect the tables that are not relevant to a given domain. Overall, the dataset contains annotations for the CEA, CTA, CPA and ITD tasks. The dataset is available on Zenodo [35].

3.3. Artifacts Availability Badge

In 2021, SemTab included a new track focusing on system usability. The main goal of this track was to mitigate a pain point in the community: the lack of publicly available, easy-to-use, and generic solution to address the needs of a variety of applications and settings. Since 2022, the usability track has been replaced by the *Artifacts Availability Badge*, that applies to both Accuracy and Datasets tracks. For SemTab 2023, the Artifact Availability Badge is awarded to DREIFLUSS [27] (Accuracy Track), TSOTSA [28] (Accuracy Track), and TSOTSATable Dataset [33] (Datasets Track).

Acknowledgements

We would like to sincerely thank all the challenge participants of this year and all the previous editions, who have all helped shape this challenge with their valuable feedback, participation at various discussions, and all their technical contributions [36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 23, 56, 21, 34, 57, 58, 59, 60, 61, 62, 63, 31]. We also thank the ISWC & OM organisers, and our sponsors. This work was also supported by the SIRIUS Centre for Scalable Data Access (Research Council of Norway), Samsung Research UK, the EPSRC projects UK FIRES and ConCur, and the HFRI project ResponsibleER (No 969). Finally, we like to acknowledge that the organization was greatly simplified by using the EasyChair conference management system and the CEUR-WS.org open-access publication service.

References

[1] G. Fan, J. Wang, Y. Li, D. Zhang, R. J. Miller, Semantics-aware dataset discovery from data lakes with contextualized column-based representation learning, Proc. VLDB Endow. 16 (2023) 1726–1739. URL: https://www.vldb.org/pvldb/vol16/p1726-fan.pdf.

[2] G. Fan, J. Wang, Y. Li, R. J. Miller, Table discovery in data lakes: State-of-the-art and future directions, in: S. Das, I. Pandis, K. S. Candan, S. Amer-Yahia (Eds.), Companion of the 2023 International Conference on Management of Data, SIGMOD/PODS 2023, Seattle, WA, USA, June 18-23, 2023, ACM, 2023, pp. 69–75. URL: https://doi.org/10.1145/3555041.3589409. doi:10.1145/3555041.3589409.

[3] A. Khatiwada, G. Fan, R. Shraga, Z. Chen, W. Gatterbauer, R. J. Miller, M. Riedewald, SANTOS: relationship-based semantic table union search, Proc. ACM Manag. Data 1 (2023) 9:1–9:25. URL: https://doi.org/10.1145/3588689. doi:10.1145/3588689.

[4] P. Ouellette, A. Sciortino, F. Nargesian, B. G. Bashardoost, E. Zhu, K. Q. Pu, R. J. Miller, RONIN: data lake exploration, Proc. VLDB Endow. 14 (2021) 2863–2866. URL: http://www.vldb.org/pvldb/vol14/p2863-nargesian.pdf. doi:10.14778/3476311.3476364.

[5] F. Nargesian, K. Q. Pu, B. G. Bashardoost, E. Zhu, R. J. Miller, Data lake organization, IEEE Trans. Knowl. Data Eng. 35 (2023) 237–250. URL: https://doi.org/10.1109/TKDE.2021.3091101. doi:10.1109/TKDE.2021.3091101.

[6] A. Khatiwada, R. Shraga, W. Gatterbauer, R. J. Miller, Integrating data lake ta-

bles, Proc. VLDB Endow. 16 (2022) 932–945. URL: https://www.vldb.org/pvldb/vol16/p932-khatiwada.pdf.

[7] A. Khatiwada, R. Shraga, R. J. Miller, DIALITE: discover, align and integrate open data tables, in: S. Das, I. Pandis, K. S. Candan, S. Amer-Yahia (Eds.), Companion of the 2023 International Conference on Management of Data, SIGMOD/PODS 2023, Seattle, WA, USA, June 18-23, 2023, ACM, 2023, pp. 187–190. URL: https://doi.org/10.1145/3555041.3589732. doi:10.1145/3555041.3589732.

[8] S. Galhotra, U. Khurana, O. Hassanzadeh, K. Srinivas, H. Samulowitz, M. Qi, Automated feature enhancement for predictive modeling using external knowledge, in: P. Papapetrou, X. Cheng, Q. He (Eds.), 2019 International Conference on Data Mining Workshops, ICDM Workshops 2019, Beijing, China, November 8-11, 2019, IEEE, 2019, pp. 1094–1097. URL: https://doi.org/10.1109/ICDMW.2019.00161. doi:10.1109/ICDMW.2019.00161.

[9] Z. Zhang, Effective and efficient semantic table interpretation using tableminer+, Semantic Web 8 (2017) 921–957.

[10] Z. Syed, T. Finin, V. Mulwad, , A. Joshi, Exploiting a Web of Semantic Data for Interpreting Tables, in: Proceedings of the Second Web Science Conference, 2010.

[11] V. Mulwad, T. Finin, A. Joshi, Automatically Generating Government Linked Data from Tables, in: Working notes of AAAI Fall Symposium on Open Government Knowledge: AI Opportunities and Challenges, 2011.

[12] V. Mulwad, T. Finin, Z. Syed, A. Joshi, T2LD: interpreting and representing tables as linked data, in: A. Polleres, H. Chen (Eds.), Proceedings of the ISWC 2010 Posters & Demonstrations Track: Collected Abstracts, Shanghai, China, November 9, 2010, volume 658 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2010. URL: https://ceur-ws.org/Vol-658/paper489.pdf.

[13] P. Buche, J. Dibie-Barthélemy, L. Ibanescu, L. Soler, Fuzzy web data tables integration guided by an ontological and terminological resource, IEEE Trans. Knowl. Data Eng. 25 (2013) 805–819. URL: https://doi.org/10.1109/TKDE.2011.245. doi:10.1109/TKDE.2011.245.

[14] G. Hignette, P. Buche, J. Dibie-Barthélemy, O. Haemmerlé, An ontology-driven annotation of data tables, in: M. Weske, M. Hacid, C. Godart (Eds.), Web Information Systems Engineering - WISE 2007 Workshops, WISE 2007 International Workshops, Nancy, France, December 3, 2007, Proceedings, volume 4832 of *Lecture Notes in Computer Science*, Springer, 2007, pp. 29–40. URL: https://doi.org/10.1007/978-3-540-77010-7_4. doi:10.1007/978-3-540-77010-7_4.

[15] G. Hignette, P. Buche, J. Dibie-Barthélemy, O. Haemmerlé, Fuzzy annotation of web data tables driven by a domain ontology, in: L. Aroyo, P. Traverso, F. Ciravegna, P. Cimiano, T. Heath, E. Hyvönen, R. Mizoguchi, E. Oren, M. Sabou, E. Simperl (Eds.), The Semantic Web: Research and Applications, 6th European Semantic Web Conference, ESWC 2009, Heraklion, Crete, Greece, May 31-June 4, 2009, Proceedings, volume 5554 of *Lecture Notes in Computer Science*, Springer, 2009, pp. 638–653. URL: https://doi.org/10.1007/978-3-642-02121-3_47. doi:10.1007/978-3-642-02121-3_47.

[16] E. Muñoz, A. Hogan, A. Mileo, Triplifying wikipedia's tables, in: A. L. Gentile, Z. Zhang, C. d'Amato, H. Paulheim (Eds.), Proceedings of the First International Workshop on Linked Data for Information Extraction (LD4IE 2013) co-located with the 12th International

Semantic Web Conference (ISWC 2013), Sydney, Australia, October 21, 2013, volume 1057 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2013. URL: https://ceur-ws.org/Vol-1057/MunozEtAl_LD4IE2013.pdf.
[17] P. Venetis, A. Y. Halevy, J. Madhavan, M. Pasca, W. Shen, F. Wu, G. Miao, C. Wu, Recovering semantics of tables on the web, Proc. VLDB Endow. 4 (2011) 528–538. URL: http://www.vldb.org/pvldb/vol4/p528-venetis.pdf. doi:10.14778/2002938.2002939.
[18] V. Efthymiou, S. Galhotra, O. Hassanzadeh, E. Jiménez-Ruiz, K. Srinivas, 1st international workshop on tabular data analysis (TaDA, in: R. Bordawekar, C. Cappiello, V. Efthymiou, L. Ehrlinger, V. Gadepally, S. Galhotra, S. Geisler, S. Groppe, L. Gruenwald, A. Y. Halevy, H. Harmouch, O. Hassanzadeh, I. F. Ilyas, E. Jiménez-Ruiz, S. Krishnan, T. Lahiri, G. Li, J. Lu, W. Mauerer, U. F. Minhas, F. Naumann, M. T. Özsu, E. K. Rezig, K. Srinivas, M. Stonebraker, S. R. Valluri, M. Vidal, H. Wang, J. Wang, Y. Wu, X. Xue, M. Zaït, K. Zeng (Eds.), Joint Proceedings of Workshops at the 49th International Conference on Very Large Data Bases (VLDB 2023), Vancouver, Canada, August 28 - September 1, 2023, volume 3462 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2023. URL: https://ceur-ws.org/Vol-3462/TADA0.pdf.
[19] E. Jimenez-Ruiz, O. Hassanzadeh, V. Efthymiou, J. Chen, K. Srinivas, SemTab 2019: Resources to Benchmark Tabular Data to Knowledge Graph Matching Systems, in: The Semantic Web: ESWC, Springer International Publishing, 2020.
[20] N. Abdelmageed, E. Jimènez-Ruiz, O. Hassanzadeh, B. König-Ries, tFood: Semantic Table Annotations Benchmark for Food Domain, 2023. URL: https://doi.org/10.5281/zenodo.7828163. doi:10.5281/zenodo.7828163.
[21] K. Korini, R. Peeters, C. Bizer, SOTAB: the WDC schema.org table annotation benchmark, in: Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2022, co-located with the 21st International Semantic Web Conference (ISWC), volume 3320 of *CEUR Workshop Proceedings*, 2022, pp. 14–19.
[22] I. Mazurek, B. Kruit, Semtab 2023 column qualifier annotation benchmark, 2023. URL: https://doi.org/10.5281/zenodo.8398347. doi:10.5281/zenodo.8398347.
[23] I. Mazurek, B. Wiewel, B. Kruit, Wikary: A Dataset of N-ary Wikipedia Tables Matched to Qualified Wikidata Statements, in: Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2022, co-located with the 21st International Semantic Web Conference (ISWC), CEUR-WS.org, 2022.
[24] E. Jiménez-Ruiz, O. Hassanzadeh, V. Efthymiou, J. Chen, K. Srinivas, V. Cutrona, Results of SemTab 2020, in: Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching co-located with the 19th International Semantic Web Conference (ISWC 2020), 2020, pp. 1–8.
[25] W. Baazouzi, M. Kachroudi, S. Faiz, Kepler-aSI at SemTab 2023, in: SemTab'23: Semantic Web Challenge on Tabular Data to Knowledge Graph Matching 2023, co-located with the 22nd International Semantic Web Conference (ISWC), 2023.
[26] S. Mehryar, R. Celebi, Semantic Annotation of Tabular Data for Machine-to-Machine Interoperability via Neuro-Symbolic Anchoring, in: SemTab'23: Semantic Web Challenge on Tabular Data to Knowledge Graph Matching 2023, co-located with the 22nd International Semantic Web Conference (ISWC), 2023.
[27] V. Parmar, A. Algergawy, DREIFLUSS: A Minimalist Approach for Table Matching, in:

SemTab'23: Semantic Web Challenge on Tabular Data to Knowledge Graph Matching 2023, co-located with the 22nd International Semantic Web Conference (ISWC), 2023.
[28] B. Foko, A. Jiomekong, H. Tapamo, J. Buisson, S. Tiwari, Exploring Naive Bayes Classifiers for Tabular Data to Knowledge Graph Matching, in: SemTab'23: Semantic Web Challenge on Tabular Data to Knowledge Graph Matching 2023, co-located with the 22nd International Semantic Web Conference (ISWC), 2023.
[29] E. G. Henriksen, A. M. Khorsid, E. Nielsen, A. M. Stück, A. S. Sørensen, O. Pelgrin, SemTex: A Hybrid Approach for Semantic Table Interpretation, in: SemTab'23: Semantic Web Challenge on Tabular Data to Knowledge Graph Matching 2023, co-located with the 22nd International Semantic Web Conference (ISWC), 2023.
[30] I. Dasoulas, D. Yang, X. Duan, A. Dimou, TorchicTab: Semantic Table Annotation with Wikidata and Language Models, in: SemTab'23: Semantic Web Challenge on Tabular Data to Knowledge Graph Matching 2023, co-located with the 22nd International Semantic Web Conference (ISWC), 2023.
[31] V.-P. Huynh, Y. Chabot, T. Labbé, J. Liu, R. Troncy., From Heuristics to Language Models: A Journey Through the Universe of Semantic Table Interpretation with DAGOBAH, in: Semantic Web Challenge on Tabular Data to Knowledge Graph Matching (SemTab), CEUR-WS.org, 2022.
[32] Y. Suhara, J. Li, Y. Li, D. Zhang, c. Demiralp, C. Chen, W.-C. Tan, Annotating columns with pre-trained language models, in: Proceedings of the 2022 International Conference on Management of Data, SIGMOD '22, Association for Computing Machinery, New York, NY, USA, 2022, p. 1493–1503. URL: https://doi.org/10.1145/3514221.3517906. doi:10.1145/3514221.3517906.
[33] A. Jiomekong, U. Melie, H. Tapamo, G. Camara, Semantic Annotation of TSOTSATable Dataset, in: SemTab'23: Semantic Web Challenge on Tabular Data to Knowledge Graph Matching 2023, co-located with the 22nd International Semantic Web Conference (ISWC), 2023.
[34] A. Jiomekong, C. Etoga, B. Foko, M. Folefac, S. Kana, V. Tsague, M. Sow, G. Camara, A large scale corpus of food composition tables, in: Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2022, co-located with the 21st International Semantic Web Conference (ISWC), CEUR-WS.org, 2022.
[35] A. Jiomekong, U. Melie, TSOTSATable dataset: a dataset of food and its composition, 2023. URL: https://doi.org/10.5281/zenodo.8169063. doi:10.5281/zenodo.8169063.
[36] S. Yumusak, Knowledge graph matching with inter-service information transfer, in: Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2020, co-located with the 19th International Semantic Web Conference (ISWC), CEUR-WS.org, 2020.
[37] G. Diallo, R. Azzi, AMALGAM: making tabular dataset explicit with knowledge graph, in: Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2020, co-located with the 19th International Semantic Web Conference (ISWC), CEUR-WS.org, 2020.
[38] W. Baazouzi, M. Kachroudi, S. Faiz, Kepler-aSI : Kepler as a Semantic Interpreter, in: Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2020, co-located with the 19th International Semantic Web Conference (ISWC),

CEUR-WS.org, 2020.
[39] N. Abdelmageed, S. Schindler, JenTab: Matching Tabular Data to Knowledge Graphs, in: Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2020, co-located with the 19th International Semantic Web Conference (ISWC), CEUR-WS.org, 2020.
[40] S. Tyagi, E. Jiménez-Ruiz, LexMa: Tabular Data to Knowledge Graph Matching using Lexical Techniques, in: Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2020, co-located with the 19th International Semantic Web Conference (ISWC), CEUR-WS.org, 2020.
[41] R. Shigapov, P. Zumstein, J. Kamlah, L. Oberländer, J. Mechnich, I. Schumm, bbw: Matching CSV to Wikidata via Meta-lookup, in: Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2020, co-located with the 19th International Semantic Web Conference (ISWC), CEUR-WS.org, 2020.
[42] M. Cremaschi, R. Avogadro, A. Barazzetti, D. Chieregato, MantisTable SE: an Efficient Approach for the Semantic Table Interpretation, in: Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2020, co-located with the 19th International Semantic Web Conference (ISWC), CEUR-WS.org, 2020.
[43] V.-P. Huynh, J. Liu, Y. Chabot, T. Labbé, P. Monnin, , R. Troncy, DAGOBAH: Enhanced Scoring Algorithms for Scalable Annotations of Tabular Data, in: Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2020, co-located with the 19th International Semantic Web Conference (ISWC), CEUR-WS.org, 2020.
[44] P. Nguyen, I. Yamada, N. Kertkeidkachorn, R. Ichise, H. Takeda, MTab4Wikidata at the SemTab 2020: Tabular Data Annotation with Wikidata, in: Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2020, co-located with the 19th International Semantic Web Conference (ISWC), CEUR-WS.org, 2020.
[45] S. Chen, A. Karaoglu, C. Negreanu, T. Ma, J.-G. Yao, J. Williams, A. Gordon, C.-Y. Lin, LinkingPark: An integrated approach for Semantic Table Interpretation, in: Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2020, co-located with the 19th International Semantic Web Conference (ISWC), CEUR-WS.org, 2020.
[46] D. Kim, H. Park, J. K. Lee, W. Kim, Generating conceptual subgraph from tabular data for knowledge graph matching, in: Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2020, co-located with the 19th International Semantic Web Conference (ISWC), CEUR-WS.org, 2020.
[47] N. Abdelmageed, S. Schindler, B. König-Ries, BiodivTab: Semantic Table Annotation Benchmark Construction, Analysis, and New Additions, in: Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2022, co-located with the 21st International Semantic Web Conference (ISWC), CEUR-WS.org, 2022.
[48] N. Abdelmageed, S. Schindler, B. König-Ries, BiodivTab: A Tabular Benchmark based on Biodiversity Research Data, in: Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2021, co-located with the 20th International Semantic Web Conference (ISWC), CEUR-WS.org, 2021.
[49] V.-P. Huynh, J. Liu, Y. Chabot, F. Deuzé, T. Labbé, P. Monnin, R. Troncy, DAGOBAH: Table

and Graph Contexts For Efficient Semantic Annotation Of Tabular Data, in: Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2021, co-located with the 20th International Semantic Web Conference (ISWC), CEUR-WS.org, 2021.

[50] W. Baazouzi, M. Kachroudi, S. Faiz, Kepler-aSI at SemTab 2021, in: Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2021, co-located with the 20th International Semantic Web Conference (ISWC), CEUR-WS.org, 2021.

[51] N. Abdelmageed, S. Schindler, JenTab Meets SemTab 2021's New Challenges, in: Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2021, co-located with the 20th International Semantic Web Conference (ISWC), CEUR-WS.org, 2021.

[52] B. Steenwinckel, F. D. Turck, F. Ongenae, MAGIC: Mining an Augmented Graph using INK, starting from a CSV, in: Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2021, co-located with the 20th International Semantic Web Conference (ISWC), CEUR-WS.org, 2021.

[53] L. Yang, S. Shen, J. Ding, J. Jin, GBMTab: A Graph-Based Method for Interpreting Semantic Table to Knowledge Graph, in: Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2021, co-located with the 20th International Semantic Web Conference (ISWC), CEUR-WS.org, 2021.

[54] P. Nguyen, I. Yamada, N. Kertkeidkachorn, R. Ichise, H. Takeda, SemTab 2021: Tabular Data Annotation with MTab Tool, in: Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2021, co-located with the 20th International Semantic Web Conference (ISWC), CEUR-WS.org, 2021.

[55] R. Avogadro, M. Cremaschi, MantisTable V: A novel and efficient approach to Semantic Table Interpretation, in: Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2021, co-located with the 20th International Semantic Web Conference (ISWC), CEUR-WS.org, 2021.

[56] M. Marzocchi, M. Cremaschi, R. Pozzi, R. Avogadro, M. Palmonari., MammoTab: a giant and comprehensive dataset for Semantic Table Interpretation, in: Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2022, co-located with the 21st International Semantic Web Conference (ISWC), CEUR-WS.org, 2022.

[57] A. Jiomekong, B. A. F. Tagne, Towards an Approach based on Knowledge Graph Refinement for Tabular Data to Knowledge Graph Matching, in: Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2022, co-located with the 21st International Semantic Web Conference (ISWC), CEUR-WS.org, 2022.

[58] W. Baazouzi, M. Kachroudi, S. Faiz, Yet Another Milestone for Kepler-aSI at SemTab 2022, in: Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2022, co-located with the 21st International Semantic Web Conference (ISWC), CEUR-WS.org, 2022.

[59] M. Cremaschi, R. Avogadro, D. Chieregato, s-elBat: a Semantic Interpretation Approach for Messy taBle-s, in: Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2022, co-located with the 21st International Semantic

Web Conference (ISWC), CEUR-WS.org, 2022.
[60] N. Abdelmageed, S. Schindler, JenTab: Do CTA solutions affect the entire scores?, in: Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2022, co-located with the 21st International Semantic Web Conference (ISWC), CEUR-WS.org, 2022.
[61] X. Li, S. Wang, W. Zhou, G. Zhang, C. Jiang, T. Hong, P. Wang, KGCODE-Tab Results for SemTab 2022, in: Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2022, co-located with the 21st International Semantic Web Conference (ISWC), CEUR-WS.org, 2022.
[62] L. Mertens, A low-resource approach to SemTab 2022, in: Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2022, co-located with the 21st International Semantic Web Conference (ISWC), CEUR-WS.org, 2022.
[63] A. Sharma, S. Dalal, S. Jain, SemInt at SemTab 2022, in: Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2022, co-located with the 21st International Semantic Web Conference (ISWC), CEUR-WS.org, 2022.